REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
11-11-2014	Final Report	17-05-2012 to 10-11-2014

4. TITLE AND SUBTITLE
Construction of a Bacterial Cell that Contains Only the Set Of Essential Genes Necessary to Impart Life

5a. CONTRACT NUMBER
HR0011-12-C-0063

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
NA

6. AUTHOR(S)
John Glass

Tony Yee

5d. PROJECT NUMBER
NA

5e. TASK NUMBER
NA

5f. WORK UNIT NUMBER
NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
J. Craig Venter Institute
9704 Medical Center Drive
Rockville, MD 20850

8. PERFORMING ORGANIZATION REPORT NUMBER
HR0011-12-C-0063.Final

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA, MTO
Prog: Living Foundries ATGC
675 N Randolph St
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NA

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
NA

14. ABSTRACT

The primary goal of this research contract was to make a minimal bacterial cell as well as establish rules for grand scale based genome design. Our intent was to construct a new strain of the bacterium, *Mycoplasma mycoides*, controlled by a genome that contains only the essential genes necessary for growth in laboratory media with a doubling time of two hours or less.

The project began with a conservative approach in which we would use information generated by gene disruption studies to make iterative, stepwise deletions. We would continue on to the next step only if the previous deletion mutant was viable and grew at a reasonable rate. While this steady approach would have ultimately proved successful, we followed a far faster and more efficient approach.

The team developed a design, build, test system for generating a minimal cell. Briefly, minimized 1/8th genome segments were designed, tested in a 7/8th wild-type background and then combined into a minimized genome. Following multiple iterations of designs for the minimal genome, we have successfully created a strain controlled by a genome that is smaller than the genome of any other known organism capable of replication in pure culture. While we expect that there are approximately 34 additional genes that can be deleted, this represents a tremendous accomplishment.

Perhaps more importantly, we have established a set of rules for genome minimization, and redesign that we believe can be applied to other bacteria to make them function for human purposes more efficiently and predictably.

15. SUBJECT TERMS
NA

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tony Yee
UU	UU	UU	UU	13	**19b. TELEPHONE NUMBER** *(include area code)* 301 795 7133

J. Craig Venter™
INSTITUTE

Construction of a Bacterial Cell that Contains Only the Set of Essential Genes Necessary to Impart Life

Report Title:	Final Technical Report
Report Number:	HR0011-12-C-0063.Final
Reporting Period:	May 17, 2012 to November 10, 2014
Contract No.:	HR0011-12-C-0063
Performing Organization:	J. Craig Venter Institute 9704 Medical Center Drive Rockville, MD 20850 USA
Author of Report:	John Glass, Anthony Yee
Principal Investigators:	Hamilton Smith, Clyde Hutchison

Abstract

The primary goal of this research contract was to make a minimal bacterial cell as well as establish rules for grand scale based genome design. Our intent was to construct a new strain of the bacterium, *Mycoplasma mycoides*, controlled by a genome that contains only the essential genes necessary for growth in laboratory media with a doubling time of two hours or less.

The project began with a conservative approach in which we would use information generated by gene disruption studies to make iterative, stepwise deletions. We would continue on to the next step only if the previous deletion mutant was viable and grew at a reasonable rate. While this steady approach would have ultimately proved successful, we followed a far faster and more efficient approach.

The team developed a design, build, test system for generating a minimal cell. Briefly, minimized 1/8th genome segments were designed, tested in a 7/8th wild-type background and then combined into a minimized genome. Following multiple iterations of designs for the minimal genome, we have successfully created a strain controlled by a genome that is smaller than the genome of any other known organism capable of replication in pure culture. While we expect that there are approximately 34 additional genes that can be deleted, this represents a tremendous accomplishment.

Perhaps more importantly, we have established a set of rules for genome minimization, and redesign that we believe can be applied to other bacteria to make them function for human purposes more efficiently and predictably.

Sponsored by
Defense Advanced Research Projects Agency
Microsystems Technology Office (MTO)
Program: Living Foundries: Advanced Tools and Capabilities for Generalizable Platforms (ATCG)
Issued by DARPA/CMO under Contract No: HR0011-12-C-0063

Table of Contents:

Summary

The goal of the project is to create a cell that contains only the set of genes that are essential for life under ideal laboratory conditions. We worked to minimize *Mycoplasma mycoides* JCVI-syn1.0 (the synthetic version of *Mycoplasma mycoides* subsp *capri*) using two approaches:

- Top Down: remove genes and clusters of genes one (or a few) at a time, proceeding only if the reduced strain is viable, with a reasonable growth rate
    - The *M. mycoides* genome has been reduced to 721,860 bp using the Tandem Repeat Endonuclease Cleavage (TREC) strategy

- Bottom Up: use the design, build, test approach to quickly reach a minimized genome
    - Design a reduced genome based on our best Tn5 gene disruption and deletion data
    - Synthesis of 1/8th genome molecules from oligonucleotides
    - Confirm the minimized 1/8th molecules support life
    - Combine all 8 minimized 1/8th segments into a genome and test for viability
        - If the genome is not viable, analyze and trouble shoot synthetic lethal effects
    - Repeat the design, build, test cycle as needed
    - We have used this approach and created a minimal cell that contains 573 kb of natural *M. mycoides* sequence.

tRNA Modularization: The initial modularization experiments are progressing. A 5.3 kb tRNA module containing the 30 tRNA genes plus the necessary promoters and terminators was constructed and sequence verified. The module was inserted into the genome in place of the largest natural cluster of tRNAs and found to be viable. In Q2, we tested the effects of removing the remaining 12 tRNA clusters from the cell individually so that substitution of each of the 30 natural tRNAs with the ones encoded in the module can be tested. All substitutions proved to be viable. Now we are resynthesizing the genome with only the 30 tRNA module.

Interspecies modules to characterize unknown genes: We observed that a gene annotated as a conserved hypothetic protein had weak similarity to biochemically characterized pseudouridine methyltransferase genes (*rlmH*) in other bacteria. We replaced essential gene MMYC_0361 with the *rlmH* gene from *Bacillus subtilis. Mycoplasma mycoides* containing the *B. subtilis rlmH* was viable. This tells us the function of a previously unknown essential gene. Efforts are now underway to do this with other *M. mycoides* essential genes of unknown function that have some similarity to characterized genes in other species.

Introduction

The goal of this research project is to build a minimal bacterial cell that contains only the genes that are required for life in ideal laboratory conditions. The pursuit of a minimized cell is critical to the advancement of biology, both as a pathway for understanding the basic requirements for cell replication and as a chassis for creating an optimized platform for any number of possible applications.

We previously reported that the *Mycoplasma mycoides* JCVI-syn1.0 genome was successfully reduced from 1078 kb to 779 kb; however, while the 779 kb genome was viable, the growth rate was far too slow to allow follow up experiments at an acceptable pace. At that point in the project, it was evident that our existing transposon data was insufficient to reach the end goal. Previously, we had only two gene categories: Essential (E), and Non-essential (N). These two categories made no provision for genes or clusters of genes that might result in slow growth phenotypes. We performed another round of transposon mutagenesis and added a third category: Impaired-growth (I). This new binning scheme informed our decisions for subsequent deletions and genome design.

The initial genome designs, "HMG" (Hail Mary Genome) and "RGD1" (Reduced Genome Design) were not viable. Interestingly, one $1/8^{th}$ segment from the HMG design and all $1/8^{th}$ segments from the RGD1 design were viable in $7/8^{th}$ wild type backgrounds. This inferred the validation of our design strategy and synthesis, and that solving the interactions between deletions in the various segments would be a major undertaking.

We addressed the interactions between deletions, termed synthetic lethality by the team, by performing additional rounds of Tn5 mutagenesis and sequencing on the various partially reduced genomes created over the duration of the project. The information gleaned from these transposon studies was used to inform our next set of designs by predicting genes switching from N to E or I as paralogous functions were removed.

Toward the completion of the contract, we had created multiple genomes within which 5 to 7 segments were reduced. The final segment that required troubleshooting was Segment 5. We deleted Cluster 33 from Segment 5, and the resulting genome (585 kb total length) was viable and contained fewer natural base pairs of natural sequence (573 kb) than any other known organism that can grow in pure culture. The genome still contains approximately 12 kb of artifactual sequence used by our group for engineering purposes; however, these bases can be deleted without consequence.

A preliminary tRNA module was designed, constructed, and introduced into the *M. mycoides* syn1.0 genome and found to be viable. Synthesis of segments with the natural tRNA loci removed is in progress.

Throughout our genome minimization process we have been establishing rules for genome remodeling. As a precursor to grand scale modularization of the genome, we have synthesized a fully modularized version of RGD1 segment 2. Those results are reported below.

Methods, Assumptions and Procedures

TOP DOWN APPROACH

The plan here was to start with the full size 1078 kb *M. mycoides* JCVI-syn1.0 synthetic genome. We have continued to use the TREC strategy to make iterative deletions in the mycoplasma genome. Targeting the N category genes and clusters proved to be effective (further discussed in the Results and Discussion section). We have made a series of strains that are progressively reduced with little to no reduction in growth rate.

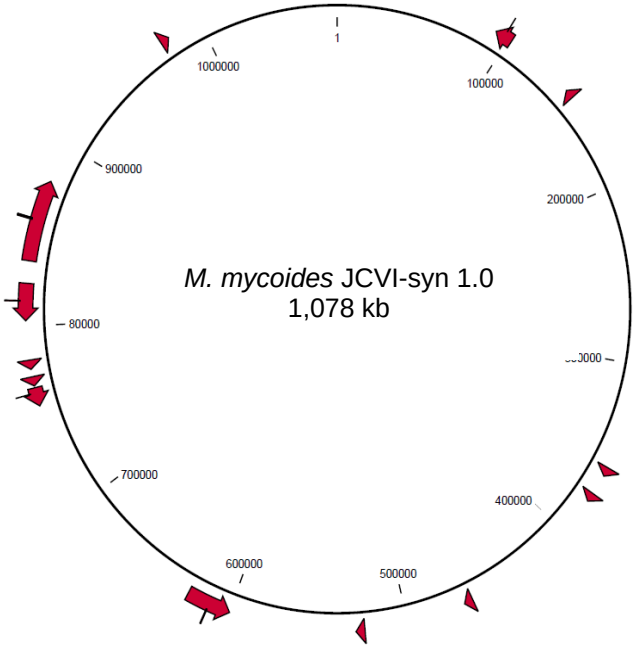
BOTTOM UP APPROACH

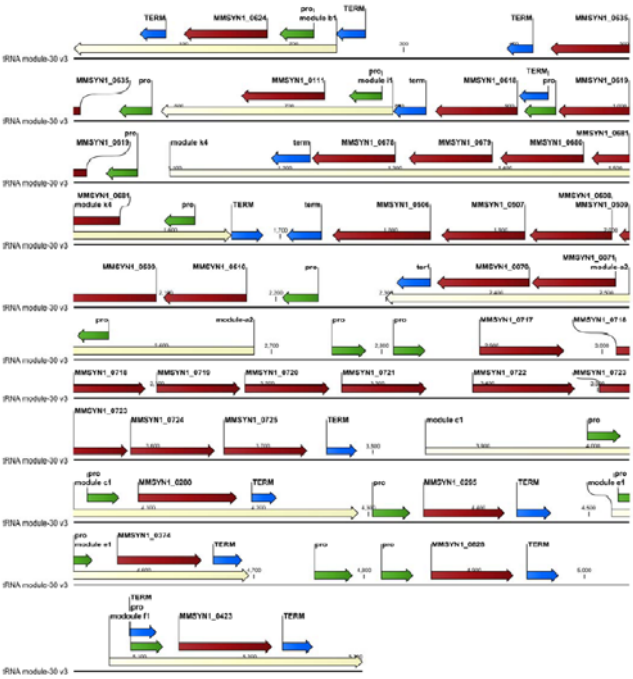
Synthesis from oligonucleotides: Multiple reduced genomes were designed, built from synthetic oligos, and tested. We used the N, E, I gene classification system to refine our design criteria and made further design updates based on Tn5 disruption of partially reduced genomes. We now have two genomes that contain all eight RGD segments.

MODULARIZATION

To test gene modularization, we have organized the 30 tRNA genes of *M. mycoides* into a single contiguous module. The module contains the coding regions, as well as the promoters and terminators needed for regulation. The tRNA genes are naturally distributed around the genome in 13 loci

Figure 1





(a) Natural distribution of tRNA genes in M. mycoides. *The tRNA gene clusters have been enlarged in Fig.1(a) to show the direction of transcription. The M. mycoides JCVI-syn1.0 genome has 8 single tRNA genes and 5 clusters of 2 to 9 genes, for a total of 30.*

(b) tRNA module design. *The 30 tRNA genes have been relocated into a single module. (Green arrows represent promoters. Red arrows show tRNA genes. Blue arrows are terminators.)*

Each of the 13 loci was synthesized by PCR using syn1.0 as the template, cloned in *E. coli* and then joined together into a single cassette with appropriate yeast markers. The cassette was inserted into syn1.0 to replace the largest cluster of 9 tRNAs at 10 o'clock on the genome map. The resulting genome is viable after transplantation. We have now made 12 other genomes in which each tRNA cluster was replaced with the synthetic 30 tRNA cluster. Each of these proved to be viable. This tells us that the tRNA module can replace all of the native tRNA genes, Based on this finding, we will now synthesize segments with the 12 other tRNA loci around the genome removed from the design.

Results and Discussion

MODULARIZATION

tRNAs: As reported in Q1, we have constructed a synthetic tRNA module that encodes all 30 tRNAs, transcriptional promoters, and terminators as a single 5.3 kb cassette. We substituted the cassette for the 9 tRNA cluster of tRNAs in *M. mycoides*, and that cell grew normally. This was an important first step. We now have a cell with a single copy of 9 tRNAs and two copies of the other 22. In Q2, we tested whether cell made previously would remain viable if individually removed the other 12 clusters of tRNAs. To date all 12 tRNA cluster deletions yielded viable *M. mycoides*; although we were surprised that some of these grow slower than the wild type cell. This exercise gives us the necessary confidence to plan to use this module rather than distributed tRNA loci in our planned fully modularized RGD (reduced genome design) genome (see below)

We are now building RGD minimal cell modules with the tRNA modules replaced by a transcriptional terminator (we do not want unintended transcription from the genes flanking the tRNA site disrupting the cell).

Module swap to characterize genes of unknown function: In another aspect of the modularization effort, we have replaced an essential gene annotated in *M. mycoides* as a conserved hypothetical protein with the *B. subtilis rlmH* gene, which produces a biochemically characterized pseudouridine methyltransferase enzyme.

Figure 2

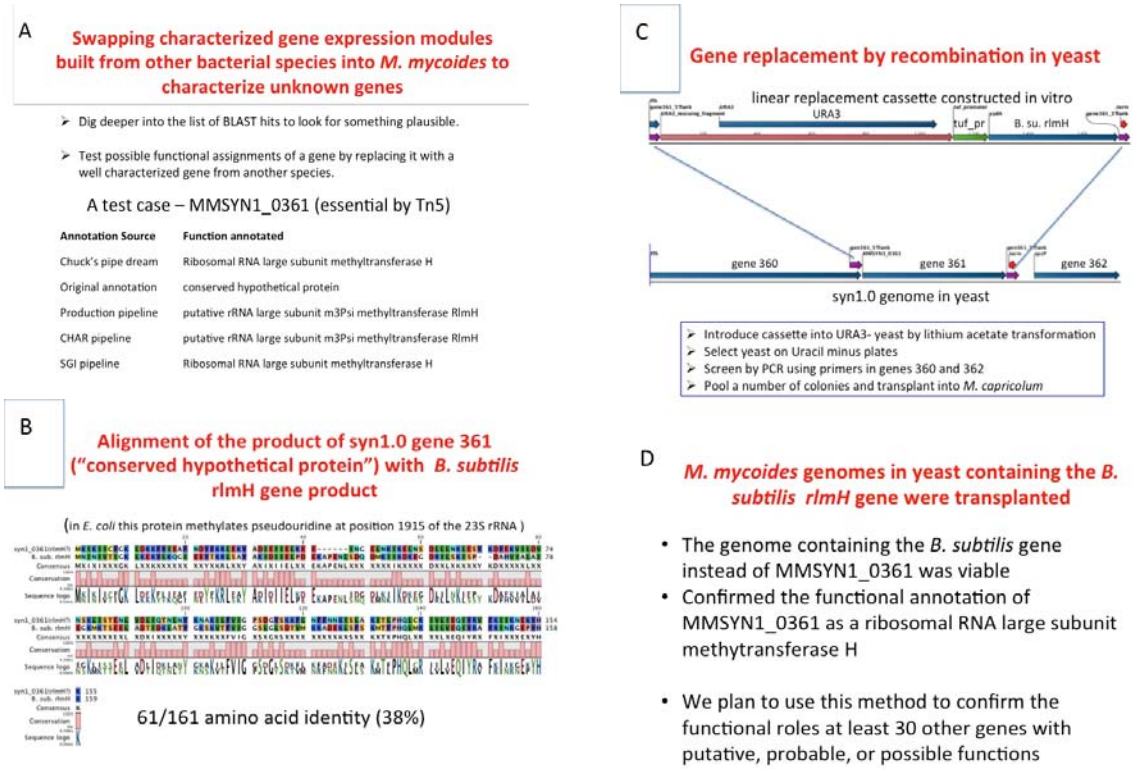


Figure 2. Swapping characterized gene expression modules built from other bacterial species into *M. mycoides* to characterize unknown genes. (A) As an example of this approach, we swapped a synthetic expression module for *Bacillus subtilis* pseudouridine methyltransferase gene rlmH with an essential gene of unknown function in M. mycoides. This characterized *rlmH* gene is slightly similar to the *M. mycoides* essential gene MMYC_0361. (B) This gene was originally annotated in *M. mycoides* as a conserved hypothetical protein. The encoded protein is 38% identical to a characterized gene in *B. subtilis* that encodes the ribosomal RNA large subunit pseudouridine methyltransferase H. (C) To swap the *B, subtilis* gene for the *M. mycoides* gene in yeast, a cassette containing a URA3 marker and the *B. subtilis rlmH* gene was constructed with the rlmH gene behind the *M. mycoides tuf* promoter, which is a strong promoter. It was exchanged into the YCp and that was transplanted into *M. capricolum*. PCRs were done using primers at the asterisks to confirm that the resulting transplants had the desired genes. (D) Analysis confirmed that *B. subtilis rlmH* could replace the *M. mycoides* MMYC_0361 gene and that the MMYC_0361 gene likely encodes ribosomal RNA large subunit methyltransferase H. We envision using this method to evaluate the function of many of the unknown genes in the minimal cell. This method can be used to evaluate other unknown bacterial genes in other organisms as well.

Functional modularization of an entire RGD segment: We have also made a more daring effort at modularization that has put us on the path to complete modularization of the RGD genome. We wrote an algorithm that completely modularized RGD segment 2. The wild type segment contains 94 genes. The RGD segment contains 41. As shown in Figure 3, the genes were re-ordered according to functions such as protein synthesis, transcription, or glycolysis. About half of the genes were just categorized as other, because there were many categories with only one gene, and there were a number of conserved hypothetical genes. When possible, genes retained their native promoter regions and transcriptional terminators. When not, they were given new promoters and terminators from those that controlled the transcription of the 53 non-essential genes left out of the RGD2 segment. Operons were broken up as needed. We did keep genes on the same DNA strand that they were on in the wild type organism. This modularized RGD2 genome segment was synthesized and combined with the other 7 wild type Syn1.0 segments, and then booted up by genome transplantation. The resulting cells had approximately the same growth rate and colony morphology as wild type Syn1.0 cells.

Figure 3

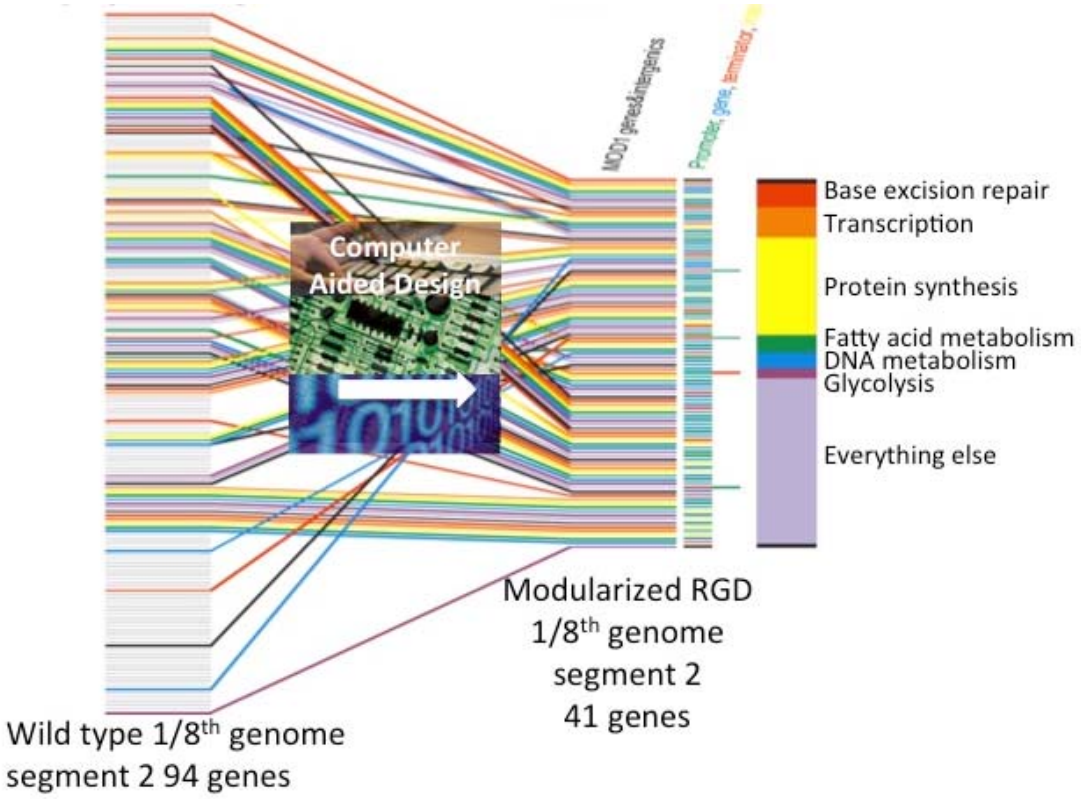


Figure 3. Cartoon of modularization of RGD segment 2.

We have now used our genome modularization software to design the other 7 modules. They are now being synthesized. Assuming each of these will work with an otherwise wild type background, we will then make a single fully modularized genome.

Fascinating Unexpected Finding about the Minimal Cell Phenotype not related to any Specific Milestone: In our efforts to characterize partially minimized *M. mycoides* strains, we performed electron microscopy. The scanning electron micrographs below are of different synthetic mycoplasma cells grown in liquid culture. The upper panels of Figure 4 show wild type *Mycoplasma mycoides* JCVI syn1.0. The bar shows 1 micron, so the cells are about ~400b nm in diameter. The cells shown in the lower panels have genomes that have the 564 Kb RGD cell or a cell with the RGD segment 6 and the other seven segments are wild type. Our analyses of cells with each individual reduced segment plus a 7/8th wild type background showed that only cells with a reduced segment 6 were giant. These RGD6 cells grow at about the same rate as wild type and produce colonies that look about the same (Figure 5). These giant cells do not appear to divide, but rather bleb or bud off daughter cells that are about the 400 nm diameter of wild type *M. mycoides* cells (Figure 6). At first we thought we could explain these enormous cells that are ~1000 X greater in volume than wild type cells by the lack of the genes encoding the cell division/cell septation proteins *ftsZ* and *ftsA* (it is also missing 78 other non-essential genes). Later we saw that a top down mutant that was essentially only missing the *ftsZ* and *ftsA* genes was the same diameter as wild type cells. We are mystified as to what are all the gene deletions that result in this phenotype. If the loss of cell division proteins FtsZ and FtsA had been sufficient to cause it, the phenotype would fit with some existing hypotheses about the evolution of cell division. Re-annotation of the whole *M.*

mycoides Syn1.0 genome resulted in predictions of the functional roles of many genes previously described only as encoding conserved hypothetical proteins. Among those are several other genes involved in cell division that are located in segment 6 (Figure 7). We now believe it was the removal of the *ftsZ* and *sepF* genes that resulted in the giant cell phenotype. We are in the process of testing this hypothesis.

Figure 4

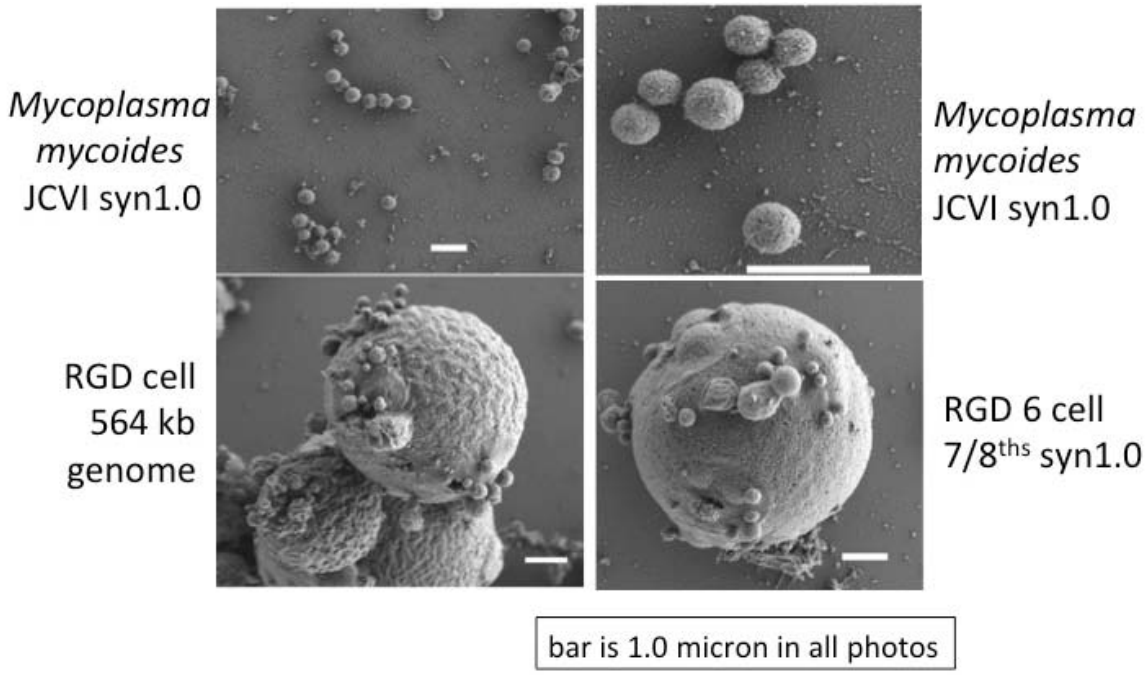


Figure 4. Electron micrographs of wild type (JCVI syn1.0) cells in the upper panels, and a partially minimized *M. mycoides* in the lower panels. The bar is 1.0 micron.

Figure 5

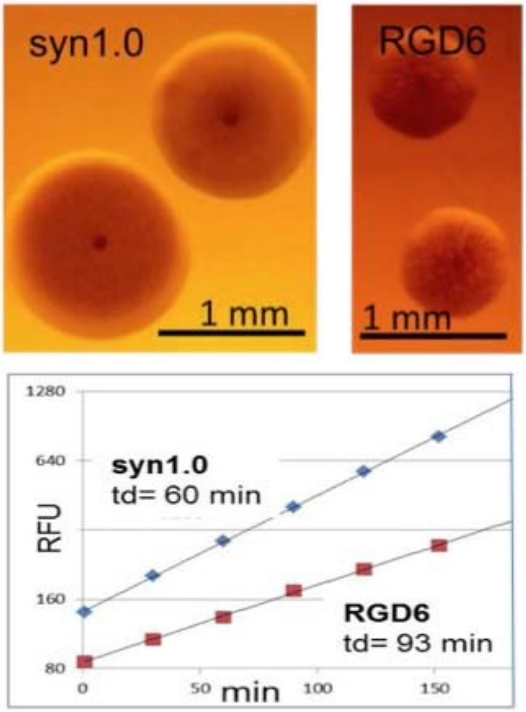


Figure 5. Giant cells grow at about the same rate as wild type cells. Colonies grown for 4 days are shown for wild type Syn1.0 cells and for a RGD6 (reduced segment 6 plus 7/8th Syn1.0) (above). Measurement of DNA accumulation in liquid cultures are shown. The slopes of the lines show the growth rates are similar, and that RGD6 grows faster than our defined minimal acceptable growth rate of one doubling every 2 hours. (below).

Figure 6

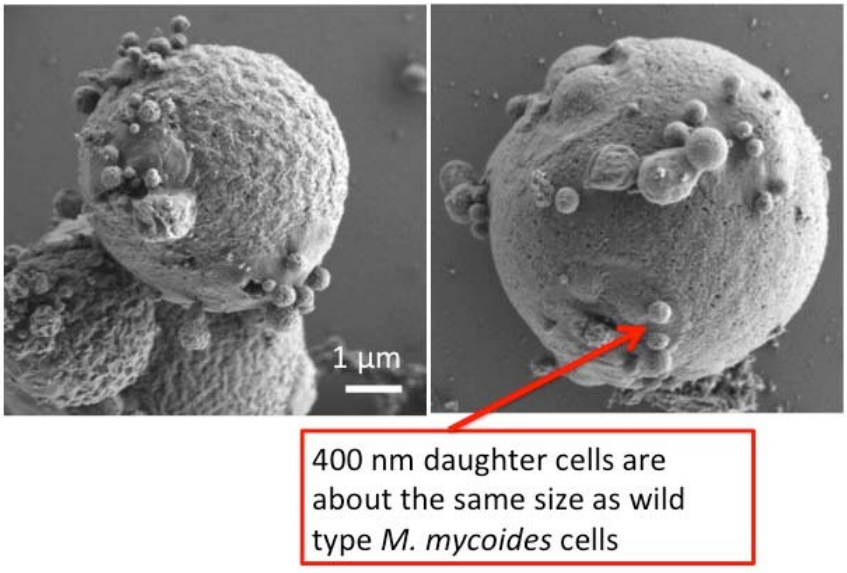


Figure 6. The giant cells appear to proliferate by budding rather than by cell division. Small near wild type 400 nm diameter blebs can be seen in the cultures and on the giant cells. We can filter the giant cells out of a culture and the small cells produce the giant cells. This method of cell proliferation is probably like the method of cell proliferation used by primordial cells before the evolution of cell division.

Figure 7

Figure 7. The cell septation/cell division gene locus of *M. mycoides* is found in segment 6. We believe it was the removal of the genes colored red in the figure that caused the giant cell phenotype. Experiments are ongoing to test this hypothesis by returning the *sepF* and *ftsZ* genes to RGD6 cells to see if that restores the normal cell phenotype.

TOP DOWN APPROACH

Iterative deletions using the TREC based approach were used to make steady progress toward a minimal genome. A table outlining the progress to date is shown below. Since the last reporting period, strain D20, D21 and D22 have been tested and found to be viable with a good qualitative growth rate (quantitative growth rate evaluation has not been performed).

Table 1

Strains	DT(min)	Genome Size (bp)	# of Genes Deleted
syn1.0	64	1,078,809	0
syn1.0D6 RE		1,062,183	17
DISs		1,048,690	31
D1		979,083	68
D2		969,069	74
D3		944,159	90
D4		931,710	97
D5		923,647	102
D6	67	908,931	108
D7		877,942	135
D8		866,271	155
D9	64	844,265	173
D10	65	828,901	181
D11		816,807	194
D12		805,506	201
D13		794,666	200
D14		784,762	207
D15		775,131	216
D16		763,995	224
D17		757,001	230
D18		749,520	235

D19		743,024	240
D20		737,508	244
D21		728,065	249
D22		721,860	255

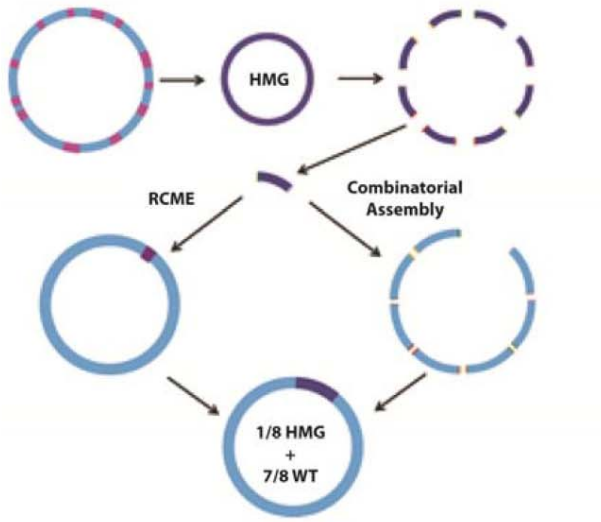
Sequential deletion of genes and clusters of genes was our original strategy, but was not the way that we ultimately pursued a minimal genome design. The outcome of performing the top down deletions is the generation of information regarding unanticipated interactions between elements of the genome that are difficult or impossible to identify using the Bottom Up approach.

BOTTOM UP APPROACH

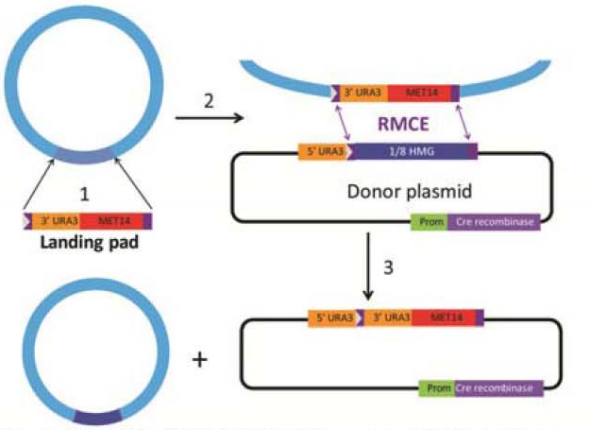
The project as proposed was based on the Top Down strategy of making iterative, step-wise deletions. Early in the project, we opted for a Bottom Up strategy whereby we would design a minimized genome, build it from oligonucleotides and test it for viability. This strategy gave us the opportunity to actually design the final genome and learn the fundamentals of genome design, rather than just arriving at a minimized genome as in the Top Down approach.

The design, build, test approach for HMG is depicted below in Figure 8. The testing of individual segments and construction into complete genomes for RGD1 & RGD2 were very similar.

Figure 8



Testing the 483 kb "Hail Mary" Genome (HMG). HMG was synthesized as 8 DNA segments with unique 200 bp overlaps (color coded) flanking each piece for genome assembly. To test the functionality or each segment, we constructed hybride genomes that are 1/8th HMG and 7/8ths wild type using Recombinase-mediated cassette exchange (RCME) or combinatorial genome assembly followed by genome transplantation.



Construction of the 1/8 HMG +7/8 wild type genome by Recombinase-mediated cassette exchange (RMCE). 1) an 1/8 genome of the M. mycoides JCVI-syn1.0 was replaced with a "landing pad" flanked by two hetero-specific lox sites; 2) a donor plasmid containing a corresponding 1/8 HMG, flanked by another two hetero-specific lox sites, was transformed into the landing pad strain; and 3) an intron-containing URA3 gene is reconstituted in the plasmid during this recombination allowing for the selection of hybrid genomes. The 1/8HMG#1, 4, and 6 were tested by this method.

The designed sizes of the RGD1 design are shown below in Table 2:

Table 2

Not I Fragment #	*M. Mycoides* JCVI-syn1.0 Length (bp)	RGD Designed Length (bp)	(RGD1)/(*M. Mycoides* JCVI-syn1.0)
1	140,739	75,732	0.54
2	120,912	49,888	0.41
3	133,208	73,958	0.56
4	131,623	82,531	0.63
5	101,708	56,501	0.56
6	189,357	80,747	0.43
7	124,976	54,482	0.44
8	137,887	66,717	0.48
Total	1,080,410	540,566	
Overlaps	-1,601	-1,601	
Genome Length	1,078,809	538,955	0.50

From the RGD1 design, all of the segments when individually combined with 7/8ths WT segments were viable. However, when the 8 RGD1 designed segments were combined into one genome in yeast, a viable cell was not obtained on transplant; we did, however, obtain viable combinations of 4 of the segments (2, 6, 7, 8). The fact that some combinations of RGD1 segments were viable was encouraging and showed that the design process was not flawed and could be used going forward. It however appeared that we would need to correct segments 1,3,4,and 5 by adding back some genes to produce a new RGD2 design.

As mentioned above, we used the N/E/I classification system in the creation of the RGD2 design. The sizes of the more conservatively designed segments are shown below:

Not I Fragment #	*M. Mycoides* JCVI-syn1.0 Length (bp)	RGD2 Designed Length (bp)	(RGD2)/(*M. Mycoides* JCVI-syn1.0)
1	140,739	90,161	0.64
2	120,912	49,888	0.41
3	133,208	88,059	0.66
4	131,623	84,750	0.64
5	101,708	61,324	0.60
6	189,357	80,747	0.43
7	124,976	54,482	0.44
8	137,887	66,717	0.48
Total	1,080,410	576,527	0.53

As of the last Quarterly Technical Report, we had created many partially reduced genomes containing between 1 and 7 reduced segments. Because there were two different genomes with 7

RGD segments and they both contained a Wild-type segment 5, we narrowed our focus to this region. We continued our work with "Clone 48", which was confirmed by sequence to be 599,897 bp.

We moved to delete Cluster 33 (14.3 kb) from segment 5 of Clone 48. We believed based on previous transposon and deletion studies that none of the genes in this cluster would result in synthetic lethal effects. The resultant genome was viable and confirmed to be the correct size by restriction analysis:

Figure 9

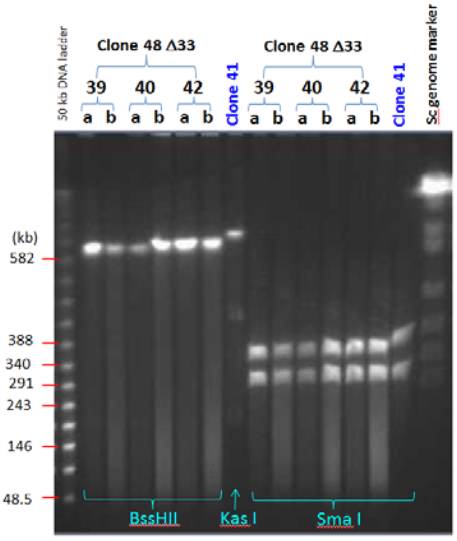


Figure 9. Restriction analysis of Cluster 33 deletion from Clone 48.

Two replicates from each of three yeast isolates (39, 40, 42) were transplanted and analyzed by restriction with BssHII and SmaI.

Digestion with BssHII produced the expected linear genome band of 588 kb.

Restriction using SmaI was expected to produce four bands, 324 kb, 263 kb, 962 bp and 16 bp. The two larger bands patterns appear as expected. (The 962 and 16 bp bands should not be visible).

The resultant genome is smaller than the genome of any other cell that can grow in axenic culture:

Clone 48 (7 RGD, plus Wild-type 5)	576,545 bp
Segment 6 Landing Pad	-2,244 bp
Yeast Centromeric Plasmid Sequence	-10,140 bp
Natural *M. mycoides* sequence	564,161 bp
M. genitalium (smallest genome of any organism that can be grown in axenic culture)	579,508 bp

Hamilton Smith predicts that there are still another 40 or so genes in the minimal genome that can be removed; however, getting below the genome size of *M. genitalium* with a reasonable growth rate was the major milestone. These remaining few genes will be removed over the next several months under other funding.

Conclusions

Tasks from the Statement of Work:

Task 1: **Complete a detailed global Tn5 transposon mutagenesis insertion map.**
Complete

Task 2: **Delete up to 27 large gene clusters**
Complete

Task 3: **Construct a preliminary modular map of the genome**
Complete

Task 4: **Make new transposon insertion map. Identify non-essential small 2-4 gene clusters. Delete small clusters.**
Complete

Task 5: **Identify non-essential single genes. Delete individual genes.**
The minimal cell is complete. Deletion of individual genes will continue under other funding.

Task 6: **Complete the removal of non-essential genes and sequences and characterize the final minimal cell product.**
We have removed all contiguous protein coding genes that are not essential for cells to double in number every 2 hours. There are 34 protein-coding genes that we plan to remove in later versions of the cell. Plus we left in both ribosomal RNA operons. While we know that only one must be retained in the genome, we know that reducing the number of ribosomal RNA operons to one results in some loss of growth rate.

Task 7: **Refine the modular map. Construction and testing of a module as a proof of principle.**
We have now shown that existing 30 tRNA genes can be combined into one module and that the tRNA genes in the module can substitute for each of the natural tRNA genes. Shortly we will have a RGD cell in which only the 30 tRNAs in the module are a source of tRNAs.

We have designed modules for glycerol metabolism, arginine hydrolysis (which would enable the cell to use arginine hydrolysis as an energy source), and amino acyl tRNA synthetases. These will be built an tested in a manner similar to what we have done for the tRNAs.

We have now fully modularized one 1/8th genome segment and found it to be viable. As a result, we have applied our modularization algorithm to the other 7 segments and are in the process of building each of those segments in modularized form. Later work will generate a fully modularized genome..

In meeting and we think exceeding all the expectations of our Living Foundries project, we have produced several important results.

First, we have built a minimal bacterial cell that we can now use as a platform to investigate the first principals of cellular life. This is something that biologists have been writing about for more than 80 years, but no one has ever actually had such a cell to do experiments with. As expected, our analyses of our RGD cell show it is comprised of genes that are widely conserved across all kingdoms of life. Figure 9 depicts that conservation of genes. To make this figure, we first ranked the RGD protein coding genes in order of annotation confidence: equivalogs, probables, putatives, possibles, and unknowns. Those genes were compared by BLASTp with genes from a number of model prokaryotes, archea, and eukaryotes. The circular diagram shows whether there is a high confidence match in other genomes with each RGD gene. The innermost circle is black because each RGD gene has a perfect match with a *M. mycoides* syn1 gene. The 15 circles are arranged based on increasing phylogenetic distance from *M. mycoides*. If a species has a gene that is a high confidence match with an RGD gene across the whole gene, then there is a line drawn for that gene. If not then there is a white space. Clearly there is a lot of gene conservation among 80% of the RGD genes. This is true even for eukaryotes (the 3 circles outside the dotted line). The unknown and possible wedges of the circle likely encode genes whose functions are largely conserved among all

life forms, but which have diverged so far from their last common ancestral gene that BLASTp will not make the connections. These findings that the minimal cell really is a sort of kernel of life encourage our efforts to use this new platform to investigate biology and also point out areas where our knowledge is especially lacking.

Second, we have developed algorithms for minimization and modularization of microbial genomes. While at present we have only used this on *Mycoplasma mycoides*, which is an organism that could only be used to generate very high value products such as pharmaceuticals, vaccines, or specialty chemicals (we think knowledge is also a very high value product), the methods will likely be applicable to other species as well. Thus one could strip away unwanted aspects of metabolism from a cell to be used for chemical production with more certainty using grand scale genome engineering, or build minimized, modularized chassis organisms.

Third, as our methods for genome manipulation continued to improve through the course of this project, we have taken *M. mycoides*, which had no genetic tools when we began our synthetic cell work in 2003, and made it the most genetically malleable organism on earth. The methods we have developed to build RGD could be adapted to work on any other bacterium we believe.

Figure 10

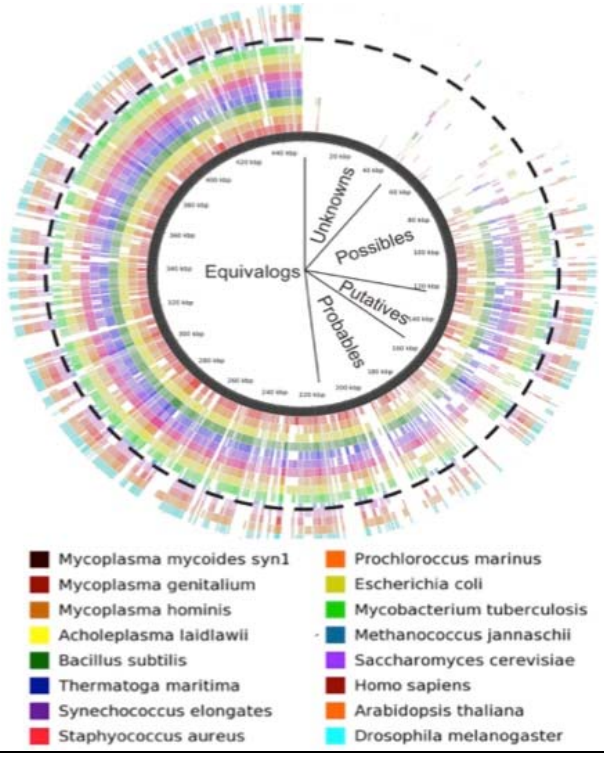


Figure 10. **BLAST Ring map of proteins remaining in RGD and homologs found in other organisms.** A BLASTp score of 1e-5 was used as the similarity cutoff. Functional classifications (unknown, possible, putative, probable, and equivalog) proceed in a stepwise fashion from no functional information (unknown) to nearly complete certainty about a genes activity (equivalog). About 28% of the genes in the reduced genome have no functional information (unknown) or an inexact “possible” activity (i.e. hydrolase, peptidase, etc.). Details for assigning a gene to a class are given in the main body of the text. Colors and organisms from top to bottom and left to right in the inset correspond to progressively larger and larger rings. White regions in the rings indicate there are no homologs to RGD in that organism. Inside the dashed circle is for prokaryotes and archea. Those outside are for eukaryotes.